

```
*****
*****
```

```
24 128 2.8 802 1 W37153 P. falciparum LSA-R-NR 5.15e-01
25 127 2.7 316 1 R26941 P. falciparum LSA gene 6.07e-01
25 127 2.7 493 1 R26944 Polypeptide fragment e 6.07e-01
27 124 2.7 567 1 W88788 Human metastasis assoc 9.95e-01
27 124 2.7 1382 1 W31867 Stenocarpus sinuatus a 2.24e+00
28 124 2.7 1382 1 W62841 CORK potassium channel 4.27e+00
29 119 2.6 1529 1 R97985 Zea mays antimicrobial 1.11e+01
30 115 2.5 33 1 W62836 Staphylococcus aureus 9.44e+00
31 109 2.4 32 110 2.4 167 1 W89759 HIV-1 protein express 9.44e+00
32 110 2.4 33 110 2.4 265 1 R12844 Human secreted protein 1.11e+01
33 109 2.4 34 109 2.4 521 1 W74802 Mannose-1-phosphate tr 8.07e+00
35 111 2.4 35 111 2.4 178 1 W30763 Escherichia coli RNase 1.51e+01
36 107 2.3 36 107 2.3 190 1 R82664 E. coli RNase E 1.51e+01
37 107 2.3 37 107 2.3 190 1 R24393 Sequence of Histidine- 2.06e+01
38 105 2.3 38 105 2.3 351 1 R24393 Human 70K UI snRNP Pro 1.29e+01
39 108 2.3 39 108 2.3 40 106 2.3 482 1 Y07657 Renal cancer associate 1.17e+01
40 106 2.3 41 106 2.3 562 1 R70491 Leucotozoan protozoa 1.77e+01
42 108 2.3 42 108 2.3 614 1 R82630 70K autoantigen, part 1.29e+01
43 104 2.2 43 104 2.2 1343 1 W31866 Mouse metastasis assoc 2.41e+01
44 103 2.2 44 103 2.2 3672 1 W31950 Human bg protein assoc 2.81e+01
45 103 2.2 3801 1 W31949 Human bg protein assoc 2.81e+01
```

```
ALIGNMENTS
```

```
*****
```

```
Release 3.1A John F. Collins, Biocomputing Research Unit.
```

```
Copyright (c) 1993-1998 University of Edinburgh, U.K.
```

```
Distribution rights by Oxford Molecular Ltd
```

```
MPSrch_PP protein - protein database search, using Smith-Waterman algorithm
```

```
Run on: Sat May 13 08:17:51 2000; MasPar time 22.33 Seconds
```

```
Tabular output not generated.
```

```
Title: >US-09-331-631-5
```

```
Description: (1-625) from US9931631.pep (1 of 4)
```

```
Perfect Score: 4631
```

```
Sequence: 1 QCMQLETSQMRRCVQSQCDK.....SPRSTKQQQPLVLVSILDFVYGF 625
```

```
Scoring table: PAM 150
```

```
Gap 11
```

```
Searched: 188963 seqs, 23686106 residues
```

```
Post-processing: Minimum Match 0% summaries
```

```
Listing first 45 summaries
```

```
Database: a-geneseq35
```

```
1:geneseqP
```

```
Statistics: Mean 36.852; Variance 168.946; scale 0.218
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4631	100.0	625	1 W62830	Macadamia integrifolia 0.00e+00	
2	4495	97.1	666	1 W62829	Macadamia integrifolia 0.00e+00	
3	4487	95.9	666	1 W62828	Macadamia integrifolia 0.00e+00	
4	1302	28.1	566	1 R20181	Sequence encoded by 67 3.01e-105	
5	1244	26.9	590	1 W62832	Gossypium hirsutum ant 6.82e-100	
6	1149	24.8	525	1 W62831	Theobroma cacao anti 3.90e-91	
7	1073	23.2	593	1 W62835	Zea mays antimicrobial 3.83e-84	
8	1041	22.5	489	1 W90341	G. max SBP2 protein. 3.33e-81	
9	992	21.4	524	1 W90339	G. max SBP1 protein. 1.04e-76	
10	972	21.0	605	1 W62838	Glycine max antimicrob 7.08e-75	
11	934	20.2	444	1 W90340	G. max truncated SBP2 2.13e-71	
12	920	19.9	626	1 W22150	Peanut allergen Ara h1 4.06e-70	
13	907	19.6	409	1 W90342	G. max truncated SBP2 6.27e-69	
14	897	19.4	614	1 W22149	Peanut allergen Ara h1 5.13e-68	
15	897	19.4	614	1 W62834	Arachis hypogaea anti 5.13e-68	
16	826	17.8	637	1 W62837	Hordeum vulgare anti 1.53e-61	
17	164	20.2	444	1 W03474	Mouse SRY-related prot 1.07e-03	
18	135	2.9	186	1 W26536	Tripanosoma cruzi anti 1.60e-01	
19	128	2.9	303	1 R60564	Dirofilaria immitis pa 1.35e-01	
20	128	2.8	432	1 W93054	Human regulatory molec 5.15e-01	
21	128	2.8	541	1 W37148	Mammalian Ena (Mena). 5.15e-01	
22	128	2.8	783	1 W37151	Mouse neural Mena+ pro 5.15e-01	
23	128	2.8	787	1 W37152	Mouse neural Mena++ pr 5.15e-01	

Db	241	LIRALKNYRLVLEANPNAVLPHTLADAILVLIGRGALKMHRDNRESYNLECGDV	300	Db	222	OHRGGDLINDORGSSGRYEEGEEHQSDNPYFDERSLSSTRFRTEEGHISVLENYGRSK	281
Qy	241	LIRALKNYRLVLEANPNAVLPHTLADAILVLIGRGALKMHRDNRESYNLECGDV	300	Qy	181	OHRGGDLMNQRGSSGRYEEGEEHQSDNPYFDERSLSSTRFRTEEGHISVLENYGRSK	240
Db	301	RIPAGTFYLINRDRNNEHLHAKFLQFOTISPGQKKEFPAGQNPPEYLSTFSKELAA	360	Db	282	LIRALKNYRLVLEANPNAVLPHTLADAILVLIGRGALKMHRDNRESYNLECGDV	341
Qy	301	RIPAGTFYLINRDRNNEHLHAKFLQFOTISPGQKKEFPAGQNPPEYLSTFSKELAA	360	Qy	241	LIRALKNYRLVLEANPNAVLPHTLADAILVLIGRGALKMHRDNRESYNLECGDV	300
Db	361	LNTQTERLRLGVIGQOREGVITRASQSOIRELTDSEBSRRMHIRREGESSSGPYLNFKR	420	Db	342	RIPAGTFYLINRDRNNEHLHAKFLQFOTISPGQKKEFPAGQNPPEYLSTFSKELAA	401
Qy	361	LNTQTERLRLGVIGQOREGVITRASQSOIRELTDSEBSRRMHIRREGESSSGPYLNFKR	420	Qy	301	RIPAGTFYLINRDRNNEHLHAKFLQFOTISPGQKKEFPAGQNPPEYLSTFSKELAA	360
Db	481	EMACPHLSGRHGRGGKKRHEEEEVHVEQPARSKREAVLWLAGHPVVVSSGENLL	540	Db	402	LNTQERLRLGVIGQOREGVITRASQSOIRELTDSEBSRRMHIRREGESSSGPYLNFKR	461
Qy	481	EMACPHLSGRHGRGGKKRHEEEEVHVEQPARSKREAVLWLAGHPVVVSSGENLL	540	Qy	361	LNTQERLRLGVIGQOREGVITRASQSOIRELTDSEBSRRMHIRREGESSSGPYLNFKR	420
Db	541	LFAGFGINAQNHNENFLAGRERNVLOQIEPOAMELAFAAFAKRKEVELFNSODESFPPGR	600	Db	462	PLYSNKYGQAYEVKPEDYQLOQDMDSVFTANTOGSMGPFFNRTRSKVVAASGEADV	521
Qy	541	LFAGFGINAQNHNENFLAGRERNVLOQIEPOAMELAFAAFAKRKEVELFNSODESFPPGR	600	Qy	421	PLYSNKYGQAYEVKPEDYQLOQDMDSVFTANTOGSMGPFFNRTRSKVVAASGEADV	480
Db	601	OHQQQSPRSTKQOQPVLVSILDFVGF	625	Db	522	EMACPHLSGRHGRGGKKRHEEEEVHVEQPARSKREAVLWLAGHPVVVSSGENLL	581
Qy	601	OHQQQSPRSTKQOQPVLVSILDFVGF	625	Qy	481	EMACPHLSGRHGRGGKKRHEEEEVHVEQPARSKREAVLWLAGHPVVVSSGENLL	540
RESULT	2			Db	582	LFAGFGINAQNHNENFLAGRERNVLOQIEPOAMELAFAAFAKRKEVELFNSODESFPPGR	641
ID	W62829	standard; Protein; 666 AA.		Db	541	LFAGFGINAQNHNENFLAGRERNVLOQIEPOAMELAFAAFAKRKEVELFNSODESFPPGR	600
AC	W62829;			Db	642	OHQQQSPRSTKQQPVLVSILDFVGF	666
DT	27-OCT-1998	(first entry)		Qy	601	OHQQQSPRSTKQQPVLVSILDFVGF	625
DE	Macadamia integrifolia antimicrobial protein.			RESULT	3		
KW	Antimicrobial Protein; Infestation; Control.			ID	R62828	standard; Protein; 666 AA.	
OS	Macadamia integrifolia.			AC	W62828;		
FR				DT	27-OCT-1998	(first entry)	
FT	Peptide	1..28		DE	Macadamia integrifolia antimicrobial protein.		
FT		/note= "signal peptide"		KW	Antimicrobial Protein; Infestation; control.		
FT	Protein	29..666		OS	Macadamia integrifolia.		
FT		/note= "mature protein"		FR			
FT	W09827805-A1.			FT	Key	Location/Qualifiers	
PN	02-JUL-1998.			FT		1..28	
PD				FT	Peptide		
PF				FT		/note= "signal peptide"	
PR	22-DEC-1997; AU00874.			FT		29..666	
PT	20-DEC-1996; AU-004275.			FT	Protein		
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.			FT		/note= "mature protein"	
PI	BOWER NI, GOULTER KC, GREEN JL, MANNERS JM, MARCUS JP;			PN	W09827805-A1.		
PT	WPI: 98-377279/32.			PN	02-JUL-1998.		
DR	N-PSDB; V42311.			PD	02-JUL-1998.		
Novel anti-microbial protein from e.g. Macadamia integrifolia - useful for controlling microbial infestations of plants or mammals				PF	22-DEC-1997; AU00874.		
PT	Claim 1; Page 39-41; 96pp; English.			PR	20-DEC-1996; AU-004275.		
PS	The sequence is that of an antimicrobial protein which can be used to control microbial infestations in plants and mammalian			PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.		
CC	Sequence 666 AA;			PI	BOWER NI, GOULTER KC, GREEN JL, MANNERS JM, MARCUS JP;		
CC	Best Local Similarity 97.1%; Score 4496; DB 1; Length 666;			DR	WPI: 98-377279/32.		
CC	Matches 604; Conservative 11; Mismatches 10; Indels 0; Gaps 0;			DR	N-PSDB; V42310.		
Db	42	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESGRQQQY	101	PT	Novel anti-microbial protein from e.g. Macadamia integrifolia - useful for controlling microbial infestations of plants or mammals		
Qy	1	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESDRQQY	60	PT	Claim 1; Page 34-36; 96pp; English.		
Db	102	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	161	PS	The sequence is that of an antimicrobial protein which can be used to control microbial infestations in plants and mammalian		
Qy	61	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	120	CC	animals.		
Db	162	OKRYEQQOREDEEKYERMKEDPKDQDPOREYECCRREQQERQYEQCQRCRQEQR	221	CC	Sequence 666 AA;		
Qy	121	OKRYEQQOREDEEKYERMKEDPKDQDPOREYECCRREQQERQYEQCQRCRQEQR	180	SO	Best Local Similarity 96.5%; Score 4487; DB 1; Length 666;		
Db	42	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESGRQQQY	101	Matches 604; Conservative 13; Mismatches 8; Indels 0; Gaps 0;			
Qy	1	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESDRQQY	60	Db	42	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESGRQQQY	101
Db	102	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	161	Qy	1	OCMOLETSGQRVCVSQCDKRFEEIDWSKVDNQDPQTDCQCQRCRQCOESDRQQY	60
Qy	61	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	120	Db	102	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	161
Db	162	OKRYEQQOREDEEKYERMKEDPKDQDPOREYECCRREQQERQYEQCQRCRQEQR	221	Qy	61	CORCKEICECEEYENRQDQOQQYEQCQRCRQHETEPRMQTCQRCRQCOESDRQQY	120

QY	96	ETEPHRM-QICQQRERRYEKEKKQKRYEQQRDEEKEEYERMEKEEDNKRDPQOREYE 154
Db	87	QCQGRGQEQDQGQRQQCQRCWQYKEBERG - HENYVHNKKRSEEEBQQRNPPY 145
AC	121	QKRYEQCORREDEEKYERMEKEEDNKRDPQOREYE 180
DB	222	OHGGGDMNPORGSSRVEEGEEFQSPNPYDERSLSTRETEEGISVLENFGSK 281
QY	181	OHGGGDMNPORGSSRVEEGEEFQSPNPYDERSLSTRETEEGISVLENFGSK 281
Db	282	LIRALKIYRLVILEANPNFLPHLDADAILLVIGGALKMTHDRRESYNELEGWV 341
QY	241	LIRALKIYRLVILEANPNFLPHLDADAILLVIGGALKMTHDRRESYNELEGWV 300
Db	342	RIPAGTIFYLINDNNERLHIAKFLQTISTPGOKEFPAGGONPEPYLTSKELA 401
QY	301	RIPAGTIFYLINDNNERLHIAKFLQTISTPGOKEFPAGGONPEPYLTSKELA 360
Db	402	LNTQTEKLKGVGQQRGSVITIASQEQLTTRDSDSRRSHWRHRRGGESSRGYNLEIKR 461
QY	361	LNTQTEKLKGVGQQRGSVITIASQEQLTTRDSDSRRSHWRHRRGGESSRGYNLEIKR 420
Db	462	PLYSNKVGQAYEVKPEYRQLQMDLSFYLTANTQSGAMGPFTNSKVVVVAASGEADV 521
QY	421	PLYSNKVGQAYEVKPEYRQLQMDLSFYLTANTQSGAMGPFTNSKVVVVAASGEADV 480
Db	522	EMACPHLSGRHGGGGKRRHEEFDVHVQVRAKSLKREAVVLAGHVVFVSSGENL 581
QY	481	EMACPHLSGRHGGGGKRRHEEFDVHVQVRAKSLKREAVVLAGHVVFVSSGENL 540
Db	582	LFAGFNGIARQNNHENFLAGGERNTLQQTEPQAMELAFAAFARKEVEESFSNSQDSIFFPRR 641
QY	541	LFAGFNGIARQNNHENFLAGGERNTLQQTEPQAMELAFAAFARKEVEESFSNSQDSIFFPRR 600
Db	642	OHQQQSRSRKQQQPVLSILDYFGF 666
QY	601	OHQQQSRSRKQQQPVLSILDYFGF 625
RESULT	4	
ID	R20181	standard; Protein; 566 AA.
AC	R20181;	
DT	16-APR-1992	(first entry)
DE	Sequence encoded by 67 kDa, cacao protein cDNA.	
KW	cocoa; flavour; vicilin; seed storage protein.	
OS	Theobroma cacao.	
PN	w09119801-A.	
PD	26-DEC-1991.	
PF	11-JUN-1991; GB-013016.	
PA	(MRC) MARS UK LTD.	
PI	Spencer ME, Hodge R, Deakin EA, Ashton S;	
DR	wpi: 92-024418/03.	
N	Recombinant cacao proteins - are responsible for flavour in cacao beans and produced in large quantities using yeast and bacterial expression vectors	
P	Claim 4; Fig 2; 59pp; English.	
S	The inventors claim a 67 kDa and 31 kDa cacao protein, and fragments, and encoding DNAs. The 47 kDa and 31 kDa proteins are derived from the 67 kDa precursor. T. cacao protein cDNA was detected in a cDNA library prepared from immature cocoa beans RNA using a probe based on the AA sequence of a CNBr peptide common to the 47 kDa and 31 kDa polypeptides. Homology searches revealed close homologies between the 67 kDa polypeptide and the vicilins, which are seed storage proteins. Sequence 566 AA;	
RESULT	5	
ID	W62832	standard; Protein; 590 AA.
AC	W62832;	
DT	27-OCT-1998	(first entry)
DE	Gossypium hirsutum antimicrobial protein.	
KW	antimicrobial protein; infestation; control.	
OS	Gossypium hirsutum.	
PN	w0927805-A1.	
PD	02-JUL-1998.	
PF	22-DEC-1997; AU0874	
PR	20-DEC-1996; AU-004275.	
PI	(RETR-) COOP RBS CEMP TROPICAL PLANT PATHOLOGY.	
DR	Bower NR, Goultier KC, Green JL, Manners JM, Marcus JP; WPI: 98-37279/32.	
PT	Novel anti-microbial protein from e.g. Macadamia integrifolia - useful for controlling microbial infestations of plants or mammals	
PS	Claim 1; Page 49-51; 96pp; English.	
CC	The sequence is that of an antimicrobial protein which can be used to control microbial infestations in plants and mammalian animals.	
SQ	Sequence 590 AA;	
Query Match	26.9%	Score 1244; DB 1; Length 590;
Best Local Similarity	39.9%	Pred. No. 6.82e-100;
Matches	225;	Conservative 149; Mismatches 143; Indels 47; Gaps 31;
CC		
CC		
CC		
SQ		
Query Match	28.1%	Score 1302; DB 1; Length 566;
Best Local Similarity	40.1%	Pred. No. 3.01e-105;
Matches	217;	Conservative 143; Mismatches 154; Indels 27; Gaps 21;
CC		
CC		
CC		
SQ		
Db	36	DPPKRYEDCKRCEWDTRGQKEQQCEESKSQYCEKDOORHREDPQRYEEQOCCR 95
QY	80	DPPKRYEDCKRCEWDTRGQKEQQCEESKSQYCEKDOORHREDPQRYEEQOCCR 129
Db	96	QOERQPOCQCRHSEQQOSROEQCQHQODRPERKQCYCRECROYEN 155
QY	130	EDEEK---YEER-NKEGDNKRDPQOREVEDCDCRICEQQEPRLO-YCQRCQCOQRH 182

Db 156 PWRRGE-REEEAEET-EEGEQEQQSHNPFFHRSFSRREHENQFRLQFASRHPIL 213
 QY :: | : | ; | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
 183 GRRGGDLMNPORGGSQRYEEGEGKQDNPYFDELSLSTRPTEBCHISLTFENFYGRSKL 242
 Db 214 RGINERFLSTILEANNTFVLPFHCDKAETKLYLTNGRGTULFLTHENKESYNTIVGGVVVK 273
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 243 RALKNYRVLVLEANDNPAVFPLTHLDADAILIVLIGRGALKMHRDNRESYNLECGDVIRI 302
 Db 274 PAGSYVYLANODNKKLITAVLHREVNNGQFEEFPAGSQRSPSYLRAFSRETELEPAIN 333
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 303 PAGTIFYLMLRDNNERLHAKFLQITSPCQYKPFPPAGSQNPPEPYLSFSKEELIAALIN 362
 Db 334 TRSECDLELFGGROSSRRQOGQMERKASOEQIRLSQAT-SPR-E- KSGE - RFAEN 387
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 363 TOTERLURGVG -Q-ORE-G-VIR-A-SQEORBLTRDSESRRWHIRRGESSRGPV 415
 Db 388 LLSQTPRYSQNSQNGRFEEACPEPFROLRDINTVTSALQNLQGSIWPHYNKATVILVTE 447
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 416 LFNRKPLPLYSKYQGQAYEVKPDYQDMSVFTANITQGSMGPFFNTRSTKVWVVAS 475
 Db 448 GNGYQEMVSHPLPROSSYBEEEDDEEDEQEEDEERRSGQYKTRRSRSRGDTWVPANT 507
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 476 GEADMEMACHL--SG-- RHGGRGKGKREEEEB--V HYEQTRARLSKREAVLWLAGH 527
 Db 508 PVTFVASYQONLRLMFGFLYQNINPDHNQRIYAGKINHV-RQWDSQAKELAFGVSSL 566
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 528 PVVFVSSGNENLLFAFGINAQN-N-H-EN-FLAGRNVLOQIEPQAMELAFAASRE 582
 Db 567 VDEFENSNQESYTS-RQRQAS 589
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 583 VEELFNSQDESIFFPPGPROHQQS 606

RESULT 6 standard; Protein: 525 AA.
 ID W62831
 AC W62831;
 DT 20-DEC-1996; AU-004275.
 PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
 TA BOWER NI, Goultier KC, Green JL, Manners JM, Marcus JP;
 DR NPI: 98-37279/32.
 PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
 PS useful for controlling microbial infestations of plants or mammals
 PS Claim 1; Page 58-60; 96pp; English.
 CC The sequence is that of an antimicrobial protein which can
 CC be used to control microbial infestations in plants and mammalian
 CC animals.
 SQ Sequence 525 AA;

Query Match 23.2%; Score 1073; DB 1; Length 593;
 Best Local Similarity 40.8%; Pred. No. 3.83e-84;
 Matches 204; Conservative 109; Mismatches 158; Indels 29; Gaps 22;

Db 25 EDNNHHHGHSQCYRRCED-PWHQPRQLEQCEEEERKQERSHEADRSGEGS 83
 QY :: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 141 EGDNNKRDQPOREYEDCRRHCEQEPRLQY-QCRRQCEQORQ--HGRGGDLMNPQG-GS 196
 Db 84 SEDEREQEKQKDRPVRPVDRSFRRVRSQSLRVLRPDEVSRLRGIRDYRAVL 143
 QY :: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 197 GRVE-EGE-EQSDN-PIYFDSLSTRFTEGHISYLENFYGRSKLRLAKNYRVL 253
 Db 144 EAMPRSFVWPSHDARCTCYVEGEGVVTBENGERSYTYRQGHFVAPAGVYIANT 203
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 254 EARPNFAVLPFLTHLDADAILIVLIGRGALKMHRDNRESYNLECGDVIRL 313
 Db 204 DGRKKLVTKLITSPCQYKPFPPAGQNPPEPYLSFSKEELIAALINOTERLWVLG 263
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 314 DNNERLHAKFLQITSPCQYKPFPPAGQNPPEPYLSFSKEELIAALINOTERLWVLG 373
 Db 264 RHGDKGKGIVRATQEQRERLHRASEGCHGPWPLPFGES RGPSLDDORPSIANQH 322
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 374 QO-RE-GVTLRASQEQRERLTDSES-R--WHIRRGESSRGPNLNPFLPYSNKYQ 428
 Db 323 QLVEADASPHDIAEHDQVSANITAGSMSAFLNFRSKTAYVNGKGAEVPHQ 382
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 429 QAYEVKPDYRQLQDMQSVF FANITQGSMGPFFNTRSTKVWVVASGEAVEMACPH-L 487
 Db 383 SGQESERERDGRSRSEEEESSEEEAEGCYHTARLSRSPGTAVVPHPEVAVS 442
 QY :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 488 S-GRHGGRG-G-GKR-----HE--EE-EEV-R-YEVARLSKREAVLWLAGHPVVFS 534

RESULT	12	Db	399	NIREGEPDLSNNGKLFLEVKPDKKNPQLDDMMUCLVETKEGALMLPHENSKAMIVVV	458	
ID	W22150; standard; Protein; 626 AA.	ID	: - :: : : : : : : : : : :	QY	415 NLFNKPLPYLSNKYQOAYEVKPD-YTROLQDMVSFTANTIQGSMGPFENTRSTKVWV	473
AC	W22150;	QY	474 ASGEADEVM-ACPHLSGRHGRGGKRHEEVEHVQEVR--AHLSKRSLAVVLAGHPV	529		
DT	29-DEC-1997 (first entry)	Db	459 NKGTCNLIELVALVARKEQQRGTEREEDEDEEEGSNSREVRYTARIKEGDVFIMPAHPV	518		
DE	Peanut allergen Ara h1.	QY	519 A-INASSE-LILIGGGINAENHR-IFLAGKRDVQDIEKQAKDIAFPSSGEGQVKLKN	576		
KW	peanut; seed storage protein; allergen; allergy; hypersensitivity; vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody; ELISA; analysis; Ara h1.	Db	530 VFVSSGNMNLILFAFGTNAQHNHENFLAGRERNVQIEQAMELAFAASRKEVELFNS	589		
KW	Araachis hypogaea strain Florunner.	QY	577 QKESHVSARQSQSOSPSPEKESP 602			
OS		QY	590 QDESIFPPGPQHQOQSRSRKQQP 615			
Key	Location/Qualifiers					
Peptide	1..22					
FT	/label= Sig_peptide					
FT	23..626					
FT	/label= Mat_protein					
FT	Modified_site					
FT	521..523					
FT	W09724139-A1.	/note= "N-glycosylation site"				
PN	10-JUL-1997.					
PD	23-SEP-1996; U15222.					
PF	04-MAR-1996; US-610424.					
PR	29-DEC-1995; US 009455.					
PA	(UYAR-) UNTV ARKANSAS.					
PT	Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;					
DR	WPI; 97-363453/33.					
DR	N-PDSB; T76613.					
PT	Peanut allergens Ara h1 and Ara h1 - used for vaccination and in two-site monoclonal antibody based ELISA					
PT	Claim 31; Page 172; 354pp; English.					
CC	This polypeptide comprises major peanut allergen Ara h1 (W22149). Its sequence was deduced from cDNA clone P41b (T76613), isolated from peanut seed cDNA using a primer (see T76616) based on an isolated Ara h1 peptide (see W4206). The sequence shows significant homology with the vicilin family of seed storage proteins of other legumes. The allergen is recognised by serum IGE from a large proportion of individuals with peanut hypersensitivity. Ara h1 and Ara h1 (see W24164) can be used to raise monoclonal antibodies which are used in a specific two-site MAB ELISA for the detection of Ara h1 or Ara h1 (claimed). IGE-binding Ara h1 antigen epitopes (see W24165-87) may be used in vaccines to protect against allergic reactions to peanut allergens, e.g. anaphylactic shock.					
CC	Sequence 626 AA;					
Query Match	19.9%; Score 920; DB 1; Length 626;					
Best Local Similarity	33.6%; Pred. No. 4.06e-70;					
Matches	190; Conservative 153; Mismatches 185; Indels 38; Gaps 26;					
Db	47 QOEPPDKIQKACESRCITKLEVPCKCVYDPQRHTGTQNQRSQPGERT-R-GPQGD-DDDR 105					
Oy	: : : : : : : : : : : : : : : : : : : : :					
78 ORDPPOQOYE-COKRCORRETEPRHOMICOGRCCRERKEKKRKKOQRYEDEQRE 135						
Db	106 QPRRERG-RWQAPGPRE-RE-RE-DWQREDWR-RPSHQOPKRIRPREGREGSQEWG 159					
Oy	: : : : : : : : : : : : : : : : :					
136 EERMKKGDNKNDPKPO-QREYEDCRRHQEQEQLQYQCORREQQEQRHGGDLNPQRG 194						
Db	160 TPGSHVR-EETSRNPNYPFPSPRFSTRYGNONGRIVRLQRTDORSOFONQNHRIVOE 218					
Oy	: : : : : : : : : : : : : : : : :					
195 GSGRYEEGEKOSDNPYFDFERSLSSTRFRTRBEGHISVLNFYGRSKLRLKRNLYLVILLE 254						
Db	219 AKPNTLVLKHDADNLVLIQOGAQATVYANGNNRKSFLDEGHALRIPSFISYIILNRH 278					
Oy	: : : : : : : : : : : : : : : : :					
255 ANPNAFLVPLTHDADAIILVLVIGGGRALKMHRDNRSYNLCFGDVIRAPGTTFLINRD 314						
Db	279 DNQNLKRAKISMVPTQGQFEDPASRDOSSVYLOPSRNLDEAFAAENEERIVILLE 338					
Oy	: : : : : : : : : : : : : : : : :					
315 NNERLHTAKFLOTISTPGQYKEFPAGQONPYSPLSFKEILEAALNTQERLRLV- 372						
Db	339 ENAGGEOEBERRGWRSTRSENNEGVVKVSKERVELTKHAKSVSKGSBEGDITNP 398					
Oy	: : : : : : : : : : : : : : : : :					
373 ---GOORE-GVI--IRASQE-Q-IRELIRD-SE-SRR-WHI-RGGE--SS-RGPY 414						
Db	152 PRHDFSEVVLNFNIKGRAVLGIVRESETEKITLEPDMHIIPIAGTPLYIVNRDENELKLLA 211					
Oy	: : : : : : : : : : : : : : : : :					
263 PTBLIDDAILVWIGGRALKMHRDNRSYNLCFGDVIRAPGTTFLINRDNNERLHIA 322						
Db	212 MHLIPVSTPGRFEFGPGSRPDPESVLSAFSNWVQAAOTPKGLERLENQNEGSIFK 271					
Oy	: : : : : : : : : : : : : : : : :					
323 KFLQTISTPGQYKEFPAGQONPYSPLSFKEILEAALNTOTERLRLGVLGQREGVIR 382						

Db 272 ISRERRVRLAPTKKSSW-WPF--GGES-KAQPNTESKRPTFSNGYGRLTVEGPDEKSL 327
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 328 ORLNLMITFNITQSMTSHNSHATKALVMGRGHQDISCPMSASDK 380
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 442 QDMDSVFANITQSMGMGPFNFNTRSTKVWVAVSGEADVEMACPHLSGRHGR 494
Qy

RESULT 14
ID W22149 standard; Protein: 614 AA.
AC W22149;
DT 29-DEC-1997 (first entry)
DE Peanut allergen Ara h1.
KW peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW vaccine; anaphylactic shock; immunotherapy; therapy;
KW monoclonal antibody; ELISA; analysis; Ara h1.
OS Arachis hypogaea strain Florunner.
KEY Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_Peptide 23..614
FT Protein /label= Mat_Protein 521..523
FT Modified_site /note= "N-glycosylation site"
PN W0924139-A1.
PD 10-JUL-1997.
PP 23-SEP-1996; U15222.
PR 04-MAR-1997; US-610424.
PA (UYAR-) UNIV ARKANS.
PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
DR N-PSDB; T76612.
PT Peanut allergens Ara h1 and Ara hII - used for vaccination and in two-site monoclonal antibody based ELISA
PT Claim 31; Page 169; 354pp; English.
CC This polypeptide comprises major Peanut allergen Ara h1 (W22149). Its sequence was deduced from cDNA clone PI7 (T76612), isolated from peanut seed cDNA using a primer (see T76616) based on an isolated Ara hI peptide (see W24206). The sequence shows significant homology with the vicilin family of seed storage proteins of other legumes. The allergen is recognised by serum IgE from a large proportion of individuals with peanut hypersensitivity. Ara hI and Ara hII (see W24164) can be used to raise monoclonal antibodies which are used in a specific two-site MAB ELISA for the detection of Ara hI or Ara hII (claimed). IgE-binding Ara hI antigen epitopes (see W24165-87) may be used in vaccines to protect against allergic reactions to peanut allergens, e.g. anaphylactic shock.

Query Match 19.4%; Score 897; DB 1; Length 614;
Best Local Similarity 33.8%; **Pred.** No. 5.13e-68;
Matches 191; **Conservative** 157; **Mismatches** 175; **Indels** 42; **Gaps** 28;

Db 45 QQEDDLKKAKACESRCKTLEYDPRCVDTGAT-NQRHPGERTRG-R-QPGDDDR-Q 100
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 101 PRREBEGG-RWGPABRE-RE-REE-DWROPREDWR-RPSHQOPRKIRPEGRGEQENG 154
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 137 ERMEKGDNKRDP-QCREYEDCCRHCQEQQPQLQCQRCQEQCQEDEEVYE 195
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 155 PGS-EVREETSRRNPFYFSPRSRSTRYGNQRVRIQRLQFQSKFQNQHNRIVOLEA 213
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 196 SGRYEEGEGEKODNDNPYFYFEDERSLSTRTEREEGHISYLLENFGRSKLRALKNYVLEA 255
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 214 RPNLVLPLKHDADNLVIQOQATVTANGNNRKNFNLDEGHALRIPSGFSTYLNHD 273
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 256 NRRAVEPLTHDADLILVIGRGAALKMHRNRESNLEGDDWIKIPIAGTTFLINRD 315
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 274 NQNLRVAKISMPPNTPGQFEDFPASSRQDQSSYLOGFSRNTLEAFNAEENIRRVLLEE 333
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 316 NERLHIAKFQLOQTISTPGQKEFFPAGGONPEPYLSTSKEILBAALNQTERLGVL-- 372
Db 334 NAGGEQEERGQRRTSRSNEQNTRSDNEGVIVKVKSEHVOELTHAKKSYSKGSSEEDITINPLR 393
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 373 --GOORE-GV1---IRASOQ--IRELTTRDPS-E-SRR-WHL-RRGES-S-RGPYNLF 417
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 394 DGEDDLSNNGFGRFEVKIDKKNFOLOQDIDMMTCVETKEGALMLPHNSKANIVVNKG 453
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 418 NKRLYLSWKYQAYKEED-YIQLOQMDVSYFLANTQGSMGPFTNTRSKVWVASY 476
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 454 TGNEELVAVRKEDQQRGRREQEWEDEESEGNSNRREVRYTARLKEGDVTFIMPAHPV 513
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 477 BADVEMACPHLSGRHGGGRGGKREHEEEBVH--QVR--ARLSKREAVVLAGHPV 529
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 514 A-INASSE-LHLLGFGINAENHRIFLAGDKVNDIEKOQDLPAGSGQVEKIKN 571
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 530 VFYSSGNNLLPAGFGINAQNHENFLAGRERNVLOQIEPQAMELAFAASRKEVEELFS 589
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 572 QRES-HFWSARPOSQSPSSPEKDQ 595
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 590 QDESIFFGCPRQIQQOSRSTKQQ 614
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

RESULT 15
ID W62834 standard; Peptide; 614 AA.
AC W62834;
DT 27-OCT-1998 (first entry)
DE Arachis hypogaea antifungal protein; infestation; control.
KW antimicrobial protein; infestation; control.
OS Arachis hypogaea.
PN W0827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874
PR 20-DEC-1996; AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bover NR, Goulier JC, Green JL, Manners JM, Marcus JP;
DR WPI; 98-37219/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia - useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 357; 96pp; English.
CC The sequence is that of an antimicrobial protein which can be used to control microbial infestations in plants and mammalian animals.
SO Sequence 614 AA;

Query Match 19.4%; Score 897; DB 1; Length 614;
Best Local Similarity 33.8%; **Pred.** No. 5.13e-68;
Matches 191; **Conservative** 157; **Mismatches** 175; **Indels** 42; **Gaps** 28;

Db 45 QQPDDLKAKACESRCKTLEYDPRCVDTGAT-NQRHPGERTRG-R-QPGDDDR-Q 100
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 101 PRREBEGG-RWGPABRE-RE-REE-DWROPREDWR-RPSHQOPRKIRPEGRGEQEWGT 154
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 137 ERMEKGDNKRDP-QCREYEDCCRHCQEQQPQLQCQRCQEQCQEDEEVYE 195
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 155 PGS-EVREETSRRNPFYFSPRSRSTRYGNQRVRIQRLQFQSKFQNQHNRIVOLEA 213
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 196 SGRYEEGEGEKODNDNPYFYFEDERSLSTRTEREEGHISYLLENFGRSKLRALKNYVLEA 255
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 214 RPNLVLPLKHDADNLVIQOQATVTANGNNRKNFNLDEGHALRIPSGFSTYLNHD 273
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 256 NRRAVEPLTHDADLILVIGRGAALKMHRNRESNLEGDDWIKIPIAGTTFLINRD 315
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 274 NÖMRVAKISMPPNTPGQFEDFPASSRQDQSSYLOGFSRNTLEAFNAEENIRRVLLEE 333
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 316 NERLHIAKFQLOQTISTPGQKEFFPAGGONPEPYLSTSKEILBAALNQTERLGVL-- 372
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 334 NAGGEQEERGQRRTSRSNEQNTRSDNEGVIVKVKSEHVOELTHAKKSYSKGSSEEDITINPLR 393
Qy |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

Mon May 15 09:51:30 2000

US-09-331-631-5-01.rag

Page 9

Qy 373 --GQQRE-GVI--IRASSEQ--IRELTRODS-E-SRR-WHI-RRGGES-S-RGYNLF 417
Db 394 DGEPPDLNNFGRLFEVVPDKKUPOLOLDMMLTCVEIKEGALMPHENSKAMVYVYNKG 453
Qy 418 NERPLYSNRYGOAYEVAPED-YRQLQMDVSIFTANTOGSAMGPFFNTRSIKVVVWASG 476
Db 454 TGNLELVAVRKEQQRREQWEEEEEDEEEEGSNREVRYTARIKEGDVFIMPAHPV 513
Qy 477 EADVEMACPHLSGRHGRGGKRHEEEEVHVE---QVR---ARLSKREAIWLAGHPV 529
Db 514 A-INASSE-LHULGGNAENIHRIFLAGDKNDIQEKAQDLAGPGSGEQVEKLKN 571
Qy 530 VEVSSGNELLFAGCINAQNHNFLAGRNLQOIEPQAMELAFAASKEVEELFNS 589
Db 572 QREBS-HFVSARPOSQSPSPEKEDQ 595
Qy 590 QDESSIFFPGPROHQQQSPRSIKQQQ 614

Search completed: Sat May 13 08:18:18 2000
Job time : 27 secs.

